A New Algorithm for Zero-Modified Models Applied to Citation Counts

Marzieh Shahmandi[1], Paul Wilson[2], and Mike Thelwall[3]

[1] *m.shahmandihounejani@wlv.ac.uk*

[2] *pauljwilson@wlv.ac.uk*

[3]m.thelwall@wlv.ac.uk

Statistical Cybermetrics Research Group, University of Wolverhampton (United Kingdom)

Abstract

Finding statistical models for citation count data is important for those seeking to understand the citing process or when using regression to identify factors that associate with citation rates. As sets of citation counts often include more or less zeros (uncited articles) than would be expected under the base distribution, it is essential to deal appropriately with them. This article proposes a new algorithm to fit zero-modified versions of discretised log-normal, hooked power-law and Weibull models to citation count data from 23 different Scopus categories from 2012. The new algorithm allows the standard errors of all parameter estimates to be calculated, and hence also confidence intervals and p-values. This algorithm can also estimate negative zero-modification parameters corresponding to zero-deflation (fewer uncited articles than expected). The results find no universal best model for the 23 categories and a given dataset may be zero-inflated relative to one model, but zero-deflated relative to another.

Introduction

It is important to identify models that fit citation distributions well for several reasons. A correct model can be used to identify anomalous sets of articles that are not fitted well by a model based upon an incorrect distribution and can help with the design of effective impact indicators. It is important when performing regression analyses to identify factors that influence citations. Also, confidence intervals for, say, model coefficients, based upon a poorly fitting model may be either too wide or too narrow, leading to incorrect estimates of possible effect sizes. It sometimes happens that the number of 0's in a given dataset are not fitted well by a distribution. This problem can often be remedied by fitting a zero-inflated or a zero-deflated (i.e. a zero-modified) distribution that allows the predicted number of zeros to approximately equal the number of zeros in a dataset.

A previous study fitted zero-inflated versions of the discretised log-normal and hooked power law distributions to citation count data from 23 Scopus categories, finding that zero-inflation occurred in the vast majority of cases (Thelwall, 2016). The zero-inflation was hypothesised to be a consequence of "inherently unciteable articles", such as magazine articles. Zero-counts due to unciteability are an example of "perfect" or "structural" zeros: data that are constrained to be zeros due to some feature of the data generating process. In contrast, other zeros are referred to as non-perfect or count zeros. In this context a non-perfect zero would be a paper that is citeable, but has not been cited. In essence, zero-inflated models seek to estimate the proportion of perfect zeros present in data, and fit a count distribution to the remaining data.

A less well-studied phenomenon is zero-deflation, where data is well-fitted by a given count distribution, but there are less zeros present in the data than would be expected under the distribution. Zero-deflation may arise for citation counts from the Web of Science (WoS), Scopus or any other citation database with selective inclusion criteria because uncited articles may be less likely to be indexed. WoS and Scopus have poorer coverage of non-English journals than of English journals so non-English journals may contribute to zero-deflation. This may be particularly relevant for fields containing nation-specific agricultural, legal, culture, or politics research.

Whilst previous studies have fitted zero inflated distributions to citation count data, none have fitted zero-deflated or zero-modified distributions to citation count data. This paper introduces zero-modified versions of the hooked power law and discretized log-normal distributions previously shown to fit citation data well (Thelwall, 2016), and also zero-modified versions of the discrete Weibull distribution. Brzezinski (2014) discusses the use of the discrete Weibull distribution to model citation counts, the discrete Weibull distribution is capable of modelling highly skewed count data with more zeros and thus is a good candidate model for citation count. Discrete Weibull distributions may be fitted to data using the R-Package DWreg (Vinciotti, 2016). The pure power law distribution is not considered because it usually requires low cited articles to be ignored for fitting and therefore is not a credible citation distribution. This paper also introduces an algorithm that fits both negative and positive zero-modification parameters, and determines the standard errors of the zero-modification (and other) parameters, which in turn enables the calculation of confidence intervals for these parameters, and the performance of statistical tests on them. The algorithms are tested on a sample set of citation data from 23 fields to assess the extent to which the new distributions fit citation count data.

Distributions

Hooked Power Law

The hooked power law (Thelwall, 2016) is a generalised version of the power law model (Pennock, David & Flake et al., 2002). The hooked power law has a probability mass function:

$$f(x;B,\alpha)=\begin{cases}A(B+x)^{-\alpha} & x=0,1,2,\dots\\ 0 & otherwise\end{cases}$$

where B and α are model parameters, and A is a constant chosen so that $\sum_{x=0}^{\infty} f(x;B,\alpha)=1$.

Discretized Log-normal

A (continuous) random variable is log-normally distributed if its logarithm is normally distributed. It has probability density function:

$$f(x;\mu,\sigma)=\frac{1}{x\,\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right),\qquad x>0,\sigma>0,\mu\in(-\infty,+\infty). \tag{1}$$

To discretise the distribution, (i.e., convert it into a form that models the situation where x is a positive integer), integrate $f(x;\mu,\sigma)$ over unit intervals about positive integer values of x, and divide by $K=\int_{0.5}^{\infty} f(x;\mu,\sigma)dx$, where f is as at (1) above. Thus, the probability mass function of the discretised log-normal distribution is:

$$g(x;\mu,\sigma)=\begin{cases}\dfrac{1}{K}\displaystyle\int_{x-0.5}^{x+0.5} f(x;\mu,\sigma)dx & x=0,1,2,3,\dots\\ 0 & otherwise\end{cases}$$

Discrete Weibull

The discrete Weibull distribution has probability mass function:

$$f(x;\mu,\sigma)=\begin{cases} q^{x^{\beta}}-q^{(x+1)^{\beta}} & x=0,1,2,3,\dots \\ 0 & otherwise \end{cases}$$

where $0<q<1$ and $\beta>0$.

Zero-modified models

A zero-modified model (see, for example, Dietz and Böhning, 2000) has the probability mass function:

$$f(x;\Theta)=\begin{cases} \omega+(1-\omega)f^{*}(x;\Theta) & x=0 \\ (1-\omega)f^{*}(x;\Theta) & x=1,2,3,\dots \\ 0 & otherwise \end{cases} \qquad (2)$$

Where Θ is a set of parameters and $f^{*}(x;\Theta)$ is a probability mass function. for negative ω the distribution is known as a zero-deflated distribution and for positive ω, it is known as a zero-inflated distribution. For $\omega=0$ the model reduces to the non-modified model, f^{*}, and if $\omega=1$ then the data would consist entirely of zeros. The zero-inflated model can be considered as a method of modelling the number of excess zeros (zero counts greater than expected under the model f^{*}), which can stem from two distinct processes, one process where zeros occur by chance, in the same manner as 1s, 2s, . . . occur; and another process by which some data are constrained to be zeros (perfect or structural zeros).

For a zero-deflated model, $\omega<0$, but may take values <-1. To see note that

$$\begin{gathered} f(0;\omega,\Theta)\geq 0 \Leftrightarrow \omega+(1-\omega)f^{*}(0;\Theta)\geq 0 \\ \Rightarrow \omega\big(1-f^{*}(0;\Theta)\big)+f^{*}(0;\Theta)\geq 0 \\ \Rightarrow \omega>\frac{f^{*}(0;\Theta)}{1-f^{*}(0;\Theta)} \end{gathered}$$

For example, if f^{*} is a Poisson distribution with parameter 0.5 then $f^{*}(0;0.5)=\exp(-0.5)=0.6065$ and hence ω is valid provided

$$\omega\geq-\frac{0.6065}{1-0.6065}=-1.54$$

The interpretation of negative values of ω is not as straightforward as those of positive values. The most straight forward interpretation is to regard $1-\omega$ as the proportionate increase in the expected number of observed positive values. For example, if $\omega=-1.5$, then we would expect to observe approximately $1-(-1.5)=2.5$ times more 1's, 2's, 3's etc in the data than we would in the non-modified model. Zero-deflation in data is usually as a consequence of some zero-counts not being included in the data. For example, Dietz and Böhning (2000) modelled zero-deflated DMFT index data from a dental epidemiological study previously published by Mendonca (1995). Specifically, the DMFT index quantifies the dental status of an individual through a count of "Decayed, Missing and Filled Teeth", and it was noted that an "incorrect sampling procedure" had led to the non-inclusion of some children whose score was zero.

Data and Methods

The data used in this article consist of citation counts for journal articles published in 2012 from 23 Scopus categories with up to 5000 journal articles for most of the categories. The citation counts to date were downloaded from Scopus in November 2017. The 5000 articles are the most recent 5000 for categories with more than 5000 articles. Whether a complete set or the most recent set of articles, this provides a coherent collection of articles with 5-6 years of citations.

A previously published algorithm fits zero-inflated discrete log-normal and zero-inflated hooked power law models to covariate free data (the zero-inflation parameter is estimated to two decimal places) (Thelwall, 2016). This model is easily extendable to zero-inflated versions of any count model, but is unable to fit negative zero-modification parameters. In this article, we include R-code for an algorithm that will enable the fitting of negative (and positive) values of ω, it also will estimate the value of ω to many decimal places and is much faster. R code to fit the models discussed in this paper is available online[1].

This algorithm is based upon maximization of the log-likelihood of the relevant zero-modified models via the *optim* command of R. The optim function offers a number of different optimisation algorithms including conjugate gradient, quasi-Newton, Nelder-Mead and simulated annealing. The default method is a derivative-free Nelder-Mead algorithm that is a method for solving high-dimensional linear optimisation problems with constraints that is not sensitive to discontinuities in the likelihood surface.

The above mentioned algorithms have advantages over techniques such as Newton-Raphson and Fisher Scoring. In particular, they optimise log-likelihood function according to the parameters simultaneously as opposed to individually, such techniques have been around a long time, but have only become practical in recent years due to improved computing power.

The model with the greatest log-likelihood usually being considered as the most appropriate model among those being considered, log likelihood does not take into account the number of parameters being estimated in the given model however.

The *optim* command has the added advantage of returning an estimate of the matrix of second order partial derivatives of the log-likelihood function, $l(f)$ corresponding to the probability mass function, f. This matrix is known as the Hessian Matrix (Faraway, 2005) of $l(f)$, and is of importance as it may be shown that the diagonal entries of its inverse are proportionate to the standard errors of the parameter estimates. This is especially useful as it enables the calculation of confidence intervals for the zero-modification parameter (as well as any other parameters), values between the interval's limits are compatible with the data, given the statistical assumptions used to compute the interval; and the performance of hypothesis tests concerning the parameters, in particular it enables test of $H_0: \omega = 0$ to determine whether there is statistical evidence of zero-modification in the data. Whilst following the publication by the American Statistical Association (Wasserstein & Lazar, 2016) of guidelines concerning the misuse of p-values and confidence interval has led to considerable debate about the use of confidence intervals and p-values, the guidelines are primarily concerned with the misuse use of p-values and confidence intervals and far from advise there abandonment. Indeed, the guidelines specifically state that "P-values can indicate how incompatible the data are with a specified statistical model''.

Several tests exist to test for zero-modification, including likelihood ratio tests, score tests, and the Wilson-Einbeck test (Wilson & Einbeck, 2018). Note that whilst the Vuong test for non-nested models has been used as a test of zero-inflation, this is erroneous (Wilson, 2015). This paper uses the Wald test (Wasserman, 2006) to test: $H_0: \omega = 0$ against the alternative: $H_1: \omega \neq 0$ with $W = \frac{\hat{\omega}}{Se(\omega)}$ where Se is the standard error of the maximum likelihood estimate of ω. We

[1] The R source code is available at https://doi.org/10.6084/m9.figshare.7643093.v1

employ the Wald test as it directly tests the significance of the estimate of the zero-modification parameter without necessitating the fitting of the non-zero-modified model.
Finally, for assessment of the fitted model, Akaike information criterion (AIC) is used to show whether one model fits the data set better than another, when the models in question contain differing numbers of parameters or predictor variables (Akaike, 1974).

Results

Proportions of uncited articles

Uncited articles are far more common is some disciplines than in others (Figure 1). Cultural Studies, Economics & Econometrics, Health Social Science, and Pharmaceutical science have the greatest proportions of zero counts respectively. The radius of the circle is proportional to the number of articles with zero-citations for each discipline. It appears that in subjects such as Pharmaceutical Science large numbers of uncited articles arise from publications that might be regarded as magazines rather than journals being included in the database.

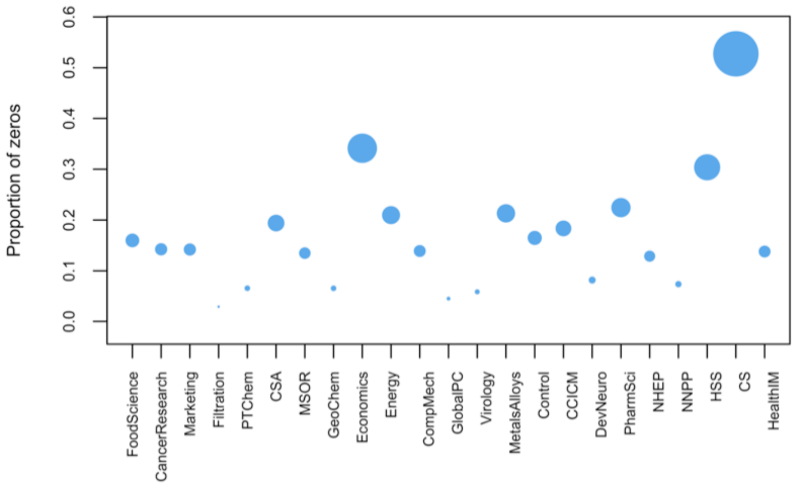

Figure 1. The proportions of uncited articles (zeros) in citation data from 23 Scopus categories.

Zero-modified discretised log-normal distribution

The zero-modification parameter estimates from the discretised log-normal distribution are all positive, the largest estimates being for Health Social Science and Economics and the smallest for Filtration & Separation and Global &Planetary Change (Figure 2, see also Table A1). The zero-inflation parameter estimates for 22 of the 23 subjects are significant at a level of significance of $\alpha=0.05$, with only Health Information Management returning a non-significant estimate. There is almost universal zero-inflation relative to the discretized log-normal distribution.

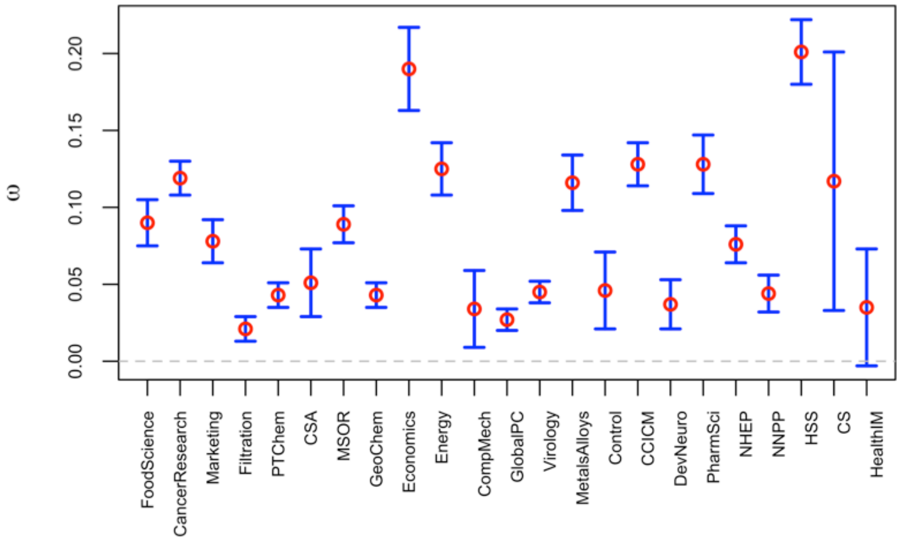

Figure 2. Zero-modification parameter parameters and 95% confidence intervals relative to a zero-modified discretised log-normal distribution for 23 Scopus categories

Zero-modified hooked power law distribution

Relative to a hooked power law distribution both significant positive (13 subjects) and significant negative (6 subjects) estimates of the zero-modification parameter occur, as well as 4 non-significant estimates (Figure 3, see also Table A2). There is both zero-inflation and zero-deflation, and possibly no zero-modification relative to the hooked power law distribution.

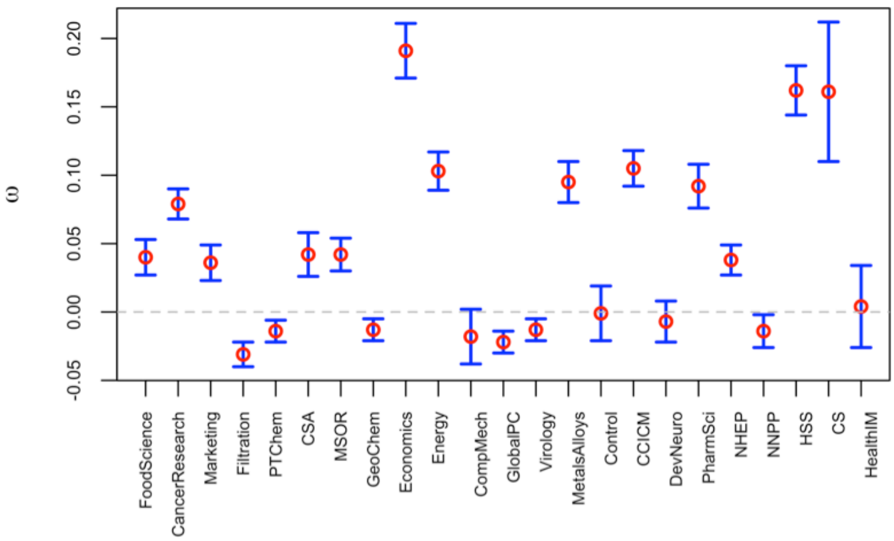

Figure 3. Zero-modification parameters and their confidence intervals relative to a zero-modified hooked power law distribution for 23 Scopus categories

Zero-modified discretised Weibull distribution

Relative to a discretise Weibull distribution only one estimate of the zero-modification parameter is significantly positive, 15 being significantly negative and 7 non-significant (Figure 4, see also Table A3). There is both zero-inflation and zero-deflation and possibly no zero-modification relative to the discretise Weibull distribution.

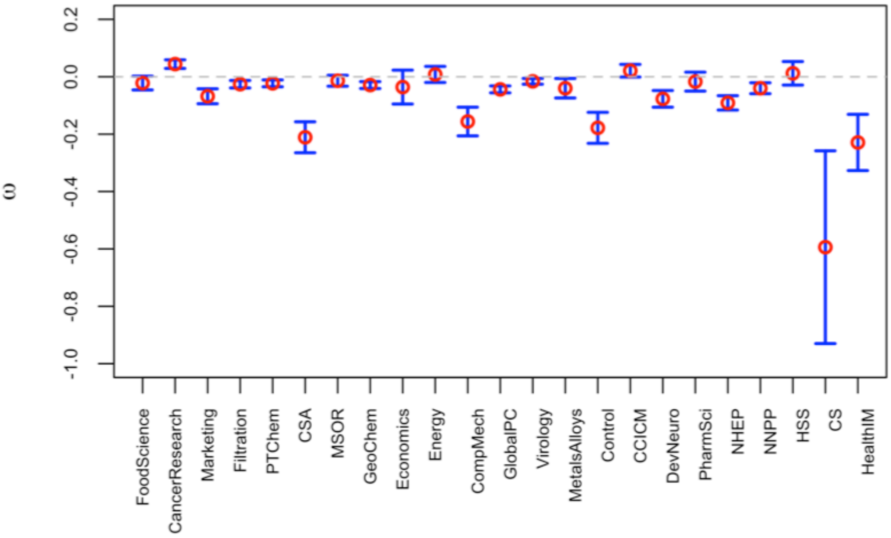

Figure 4. Zero-modification parameters and their confidence intervals relative to a zero-modified discretise Weibull distribution for 23 Scopus categories

Discussion

The results comparing distributions are limited by considering only one year and a small sample of Scopus categories. Other years and results may well give different results. The citation count distributions may also be affected by articles published in January having almost a year longer to be cited than articles published in December. Moreover, the analysis does not take into account factors that influence citation counts, such as individual, institutional and international collaboration; journal and reference impacts; abstract readability; reference and keyword totals; paper, abstract and title lengths.

It is clear from the results that zero-modification occurs relative to a given distribution. For example, the estimated value of the zero-inflation parameter for Neuropsychology and Physiological Psychology is 0.044 relative to the discretised log-normal distribution, but −0.040 relative to a hooked power law distribution, both estimates being significant. Thus, with the former distribution as the base-model there is statistical evidence of zero-inflation and hence "unciteable articles" within the field, but with the latter as the base distribution there is no such evidence of unciteable articles; instead there is evidence that some uncited articles may have been excluded. It is thus important to determine the best fitting base distribution to accurately determine the presence of zero-inflation of zero-deflation (or the absence of either), and the presence of zero inflation/deflation for one model is insufficient to prove that there are perfect or omitted zeros.

The zero-modified hooked power law distribution is the best fitting model for 13 subject areas, the zero-modified Weibull fitting best for 6 subject areas, the other 4 being best fitted by the zero-modified discrete log-normal (Table 1). Thus, whilst (as in Thelwall, 2016), the zero-modified hook power law distribution is still the best fitting model in the majority of cases, the zero-modified discrete Weibull distribution is also a good candidate in some cases. This indicates that either citation counts are best modelled by a single universal distribution that has not yet been considered, or that it is inadvisable to attempt to model citation counts without incorporating predictors, for example, number of authors, into the analyses.

Table 1. AIC values for zero-modified and standard versions of discretised log-normal, hooked power law and Weibull for 23 Scopus categories. Best fitting distributions are in bold.

Subjects	AIC ZMDLN	AIC ZMHPL	AIC ZMWeibull	AIC DLN	AIC HPL	AIC Weibull
Food Science	31448.58	**31393.14**	31398.72	31556.66	31428.84	**31402.40**
Cancer Research	37461.12	**37425.02**	37534.54	37914.00	37687.96	**37564.80**
Marketing	32366.34	**32348.92**	32451.12	32462.34	**32381.32**	32484.36
Filtration & Separation	**13469.50**	13535.42	13553.32	**13507.82**	13566.04	13569.94
Physical & Theoretical Chemistry	36176.02	**36155.10**	36200.76	36304.14	**36166.10**	36216.52
Computer Science Application	**31554.04**	31562.34	31594.68	**31571.44**	31585.56	31707.74
Management Science & Operations Research	34599.58	**34576.70**	34634.12	34771.90	**34632.34**	34636.38
GeoChemistry & Petrology	36667.80	**36661.82**	36738.60	36789.84	**36670.72**	36762.96
Economics & Econometrics	26953.04	26970.90	**26933.90**	27067.86	27195.56	**26935.50**
Energy Engineering & Power Technology	31841.10	31813.28	**31748.70**	31982.44	32000.72	**31749.06**
Computational Mechanics	15564.98	**15559.42**	15591.68	15571.52	**15562.38**	15654.46
Global & Planetary Change	**29959.92**	29986.90	30089.84	30019.14	**30012.50**	30145.58
Virology	**37266.22**	37369.72	37448.60	37471.16	**37380.06**	37457.88
Metals & Alloys	31544.04	31526.70	**31512.86**	31650.52	31671.40	**31519.04**
Control & Optimization	18375.46	**18369.88**	18420.50	18386.84	**18369.90**	18493.82
Critical Care & Intensive Care Medicine	35467.04	35432.32	**35428.70**	35706.52	35726.20	**35431.84**
Developmental Neuroscience	14550.92	**14526.48**	14579.18	14570.48	**14527.44**	14615.60
Pharmaceutical Science	29745.64	29721.02	**29715.20**	29864.16	29844.84	**29716.36**
Nuclear & High Energy Physics	35596.62	**35554.86**	35785.94	35716.84	**35600.62**	35856.58
Neuropsychology & Physiological Psych	20236.10	**20219.00**	20266.78	20289.52	**20224.16**	20287.24
Health Social Science	27138.34	**27117.98**	27183.22	27324.98	27372.70	**27183.54**
Cultural Studies	16745.46	16750.88	**16744.14**	**16751.40**	16774.10	16777.56
Health Information Management	6772.300	**6768.460**	6811.520	6775.280	**6768.540**	6854.220

Conclusion

This article introduced the zero-modified hooked power law, discrete log-normal and Weibull distributions. The new method also allows the estimation of both positive and negative zero-modification parameters, enabling the determination of confidence intervals for and statistical tests of parameter estimates. The results showed that each distribution fits citation count data better than the others for some Scopus categories. The results also show that both zero-inflation and zero-deflation occur for citation count data, but changing a base model can alter one type to another. As a consequence of this, it is important to be wary of making definitive statements concerning zero-inflation or zero-deflation.

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6):716–723.

Brzezinski, M. (2015). Power laws in citation distributions: evidence from scopus. Scientometrics, 103(1):213–228.

Dietz, E. & Böhning, D. (2000). On estimation of the Poisson parameter in zero-modified Poisson models. Computational Statistics & Data Analysis, 34(4):441 – 459.

Faraway, J. J. (2005). Extending the linear model with r. generalized linear, mixed effects and nonparametric regression models. Chapmen Hall, Boca Raton, FL, USA, ix.

Greenland, S., Senn, S., Rothman, K., Carlin, J., Poole, C., Goodman, S., & Altman, D. (2016). Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. European Journal of Epidemiology, 31(4):337–350.

Mendonca, L. (1995). Longitudinalstudie zu kariespraventiven methoden, durchgefuhrt bei 7 bis 10 jahrigen urbanen kindern in belo horizonte (brasilien). Inaugural-Dissertation zur Erlangung der zahnmedizinischen Doktorwurde am Fachbereich Zahn, Mund und Kieferheilkunde der Freien Universitat Berlin.

Pennock, D. M., Flake, G. W., Lawrence, S., Glover, E. J., & Giles, C. L. (2002). Winners don't take all: Characterizing the competition for links on the web. PNAS, 99(8):5207–5211.

Thelwall, M. (2016). Are there too many uncited articles? zero-inflated variants of the discretised lognormal and hooked power law distributions. Journal of Informetrics, 10(2):622 – 633.

Vinciotti, V. (2016). DWreg: Parametric Regression for Discrete Response. R package version 2.0.

Wasserman, L. (2006). All of Nonparametric Statistics (Springer Texts in Statistics). Springer-Verlag, Berlin, Heidelberg.

Wilson, P. (2015). The misuse of the Vuong test for non-nested models to test for zero-inflation. Economics Letters, 127(C):51–53.

Wilson, P. & Einbeck, J. (2018). A new and intuitive test for zero-modification. Statistical Modelling. Available Online.https://doi.org/10.1177/1471082X18762277.

Wasserstein, R. L. and Lazar, N. A. (2016). The asa's statement on p-values: Context, process, and purpose. The American Statistician, 70(2):129–133.

Appendix: Extra details of the parameter estimates.

Table A1. Parameter estimates: Zero-modified discretised log-normal for 23 Scopus categories

Subjects	ω	SE_{ω}	CIL_{ω}	CLR_{ω}	p-value	μ	σ
Food Science	0.090	0.0075	0.075	0.105	0.00000	1.82	1.02
Cancer Research	0.119	0.0054	0.108	0.13	0.00000	2.43	1.06
Marketing	0.078	0.0071	0.064	0.092	0.00000	1.98	1.10
Filtration & Separation	0.021	0.0042	0.013	0.029	0.00000	2.58	0.90
Physical & Theoretical Chemistry	0.043	0.0040	0.035	0.051	0.00000	2.28	0.94
Computer Science Application	0.051	0.0114	0.029	0.073	0.00001	1.56	1.29
Management Science & Operations Research	0.089	0.0061	0.077	0.101	0.00000	2.10	1.05
GeoChemistry & Petrology	0.043	0.0040	0.035	0.051	0.00000	2.31	0.96
Economics & Econometrics	0.190	0.0136	0.163	0.217	0.00000	1.34	1.29
Energy Engineering & Power Technology	0.125	0.0087	0.108	0.142	0.00000	1.82	1.15
Computational Mechanics	0.034	0.0126	0.009	0.059	0.00697	1.65	1.06
Global & Planetary Change	0.027	0.0037	0.020	0.034	0.00000	2.49	1.00
Virology	0.045	0.0035	0.038	0.052	0.00000	2.41	0.91
Metals & Alloys	0.116	0.0094	0.098	0.134	0.00000	1.75	1.18
Control & Optimization	0.046	0.0129	0.021	0.071	0.00036	1.57	1.08
Critical Care & Intensive Care Medicine	0.128	0.0070	0.114	0.142	0.00000	2.16	1.20
Developmental Neuroscience	0.037	0.0081	0.021	0.053	0.00000	2.18	1.07
Pharmaceutical Science	0.128	0.0096	0.109	0.147	0.00000	1.66	1.09
Nuclear & High Energy Physics	0.076	0.0063	0.064	0.088	0.00000	2.13	1.13
Neuropsychology & Physiological Psych	0.044	0.0059	0.032	0.056	0.00000	2.21	0.98
Health Social Science	0.201	0.0108	0.180	0.222	0.00000	1.54	1.07
Cultural Studies	0.117	0.0431	0.033	0.201	0.00664	0.10	1.20
Health Information Management	0.035	0.0193	-0.003	0.073	0.06976	1.83	1.24

Table A2. Parameter estimates: Zero-modified hooked power law for 23 Scopus categories

Subjects	ω	SE_{ω}	CIL_{ω}	CLR_{ω}	p-value	μ	σ
Food Science	0.040	0.0067	0.027	0.053	0.00000	40.21	6.37
Cancer Research	0.079	0.0055	0.068	0.090	0.00000	61.35	5.34
Marketing	0.036	0.0064	0.023	0.049	0.00000	27.28	4.18
Filtration & Separation	-0.031	0.0048	-0.040	-0.022	0.00000	190.92	12.52
Physical & Theoretical Chemistry	-0.014	0.0042	-0.022	-0.006	0.00086	121.18	10.90
Computer Science Application	0.042	0.0084	0.026	0.058	0.00000	11.31	2.96
Management Science & Operations Research	0.042	0.0059	0.030	0.054	0.00000	40.78	5.15
GeoChemistry & Petrology	-0.013	0.0042	-0.021	-0.005	0.00197	93.37	8.47
Economics & Econometrics	0.191	0.0101	0.171	0.211	0.00000	10.05	3.06
Energy Engineering & Power Technology	0.103	0.0073	0.089	0.117	0.00000	26.56	4.38
Computational Mechanics	-0.018	0.0103	-0.038	0.002	0.08054	20.28	4.39
Global & Planetary Change	-0.022	0.0039	-0.030	-0.014	0.00000	79.88	6.37
Virology	-0.013	0.0039	-0.021	-0.005	0.00086	127.17	10.26
Metals & Alloys	0.095	0.0076	0.080	0.110	0.00000	19.76	3.77
Control & Optimization	-0.001	0.0102	-0.021	0.019	0.92190	17.34	4.14
Critical Care & Intensive Care Medicine	0.105	0.0064	0.092	0.118	0.00000	31.22	3.86
Developmental Neuroscience	-0.007	0.0075	-0.022	0.008	0.35065	45.73	5.21
Pharmaceutical Science	0.092	0.0080	0.076	0.108	0.00000	23.23	4.66
Nuclear & High Energy Physics	0.038	0.0058	0.027	0.049	0.00000	31.67	4.15
Neuropsychology & Physiological Psych	-0.014	0.0060	-0.026	-0.002	0.01963	70.08	7.30
Health Social Science	0.162	0.0091	0.144	0.180	0.00000	17.89	4.33
Cultural Studies	0.161	0.0260	0.110	0.212	0.00000	4.25	3.47
Health Information Management	0.004	0.0155	-0.026	0.034	0.79636	13.47	2.94

Table A3. Parameter estimates: Zero-modified Weibull for 23 Scopus categories

Subjects	ω	SE_ω	CIL_ω	CLR_ω	p-value	μ	σ
Food Science	-0.022	0.0121	-0.046	0.002	0.06904	0.82	0.80
Cancer Research	0.044	0.0076	0.029	0.059	0.00000	0.90	0.81
Marketing	-0.068	0.0133	-0.094	-0.042	0.00000	0.80	0.70
Filtration & Separation	-0.026	0.0064	-0.039	-0.013	0.00005	0.95	1.00
Physical & Theoretical Chemistry	-0.023	0.0059	-0.035	-0.011	0.00010	0.91	0.94
Computer Science Application	-0.211	0.0273	-0.265	-0.157	0.00000	0.67	0.56
Management Science & Operations Research	-0.014	0.0098	-0.033	0.005	0.15313	0.85	0.78
GeoChemistry & Petrology	-0.029	0.0062	-0.041	-0.017	0.00000	0.91	0.90
Economics & Econometrics	-0.036	0.0303	-0.095	0.023	0.23479	0.64	0.57
Energy Engineering & Power Technology	0.008	0.0142	-0.020	0.036	0.57318	0.80	0.72
Computational Mechanics	-0.156	0.0256	-0.206	-0.106	0.00000	0.74	0.70
Global & Planetary Change	-0.044	0.0063	-0.056	-0.032	0.00000	0.91	0.86
Virology	-0.016	0.0052	-0.026	-0.006	0.00209	0.93	0.95
Metals & Alloys	-0.040	0.0172	-0.074	-0.006	0.02004	0.76	0.66
Control & Optimization	-0.178	0.0275	-0.232	-0.124	0.00000	0.71	0.66
Critical Care & Intensive Care Medicine	0.021	0.0113	-0.001	0.043	0.06311	0.83	0.70
Developmental Neuroscience	-0.077	0.0146	-0.106	-0.048	0.00000	0.85	0.76
Pharmaceutical Science	-0.017	0.0167	-0.050	0.016	0.30870	0.76	0.72
Nuclear & High Energy Physics	-0.091	0.0126	-0.116	-0.066	0.00000	0.80	0.66
Neuropsychology & Physiological Psych	-0.040	0.0095	-0.059	-0.021	0.00003	0.89	0.87
Health Social Science	0.012	0.0211	-0.029	0.053	0.56955	0.70	0.67
Cultural Studies	-0.594	0.1712	-0.930	-0.258	0.00052	0.30	0.49
Health Information Management	-0.229	0.0501	-0.327	-0.131	0.00000	0.70	0.56